Spatial Variability and Robust Interpolation of Seafloor Sediment Properties Using the SEABED Data Bases

John A. Goff
University of Texas Institute for Geophysics
4412 Spicewood Springs Rd., Building 600
Austin, TX 78759
Phone: 512-471-047 Fax: 512-471-0999 E-mail: goff@ig.utexas.edu

LONG-TERM GOALS

This project is a collaborative effort with C. Jenkins at the Univ. Colorado. The long-term goals are to:

(1) advance the understanding of the spatial variability of seabed properties as a function of geologic environment;

(2) develop robust means of interpolation in the presence of uncertain data,

(3) provide for the estimation of uncertainty in the interpolation at unsampled locations, and enable investigation of optimal survey design to minimize uncertainties; and

(4) publish a computational/database structure capable of producing seafloor maps of wide geographic extent, for multidisciplinary use - in global change issues, defense, engineering, and ecology.

OBJECTIVES

There are huge amounts of data that describe the character of the seabed based on samplings and direct inspections made over decades. That data continues to grow rapidly and are still the richest and most detailed source of seabed information, an essential adjunct of modern remote sensed data types. One impediment to the wider use of the sample data is that it is difficult to draw area-maps (grid or polygon) from it. Other impediments such as word-based data and 3D-stratigraphic issues have seen much progress recently. This project aims to solve the remaining problem of reliable map generation. At project end the marine community will have a toolbox of interpolation tools for surficial seabed mapping. Researchers and operational groups will then be able to input detailed, spatially varying values on seafloor properties into models to increase the accuracy of sediment transport and acoustic propagation predictions.

The issue of seabed variability will also be investigated. Very little is known or published on this topic, primarily because a large, comprehensive data base has not heretofore been available. The SEABED data bases represent a tremendous opportunity. Our objective here will be to determine seabed variability properties as a function of environment; i.e., water depth, geology, sediment type, oceanography, etc. This in itself will constitute a significant scientific contribution. Such knowledge could provide a basis for predicting the structure of seabed variability in undersampled regions.

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **30 SEP 2005**	2. REPORT TYPE	3. DATES COVERED **00-00-2005 to 00-00-2005**
4. TITLE AND SUBTITLE **Spatial Variability and Robust Interpolation of Seafloor Sediment Properties Using the SEABED Data Bases**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Texas Institute for Geophysics,4412 Spicewood Springs Rd., Building 600,Austin,TX,78759**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
code 1 only

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**9**	

Variability will also form a primary constraint in the investigation of robust interpolation techniques, and for optimal survey design decision making.

APPROACH

Goff's primary contribution to this project will be in statistical analysis of grain size data and developing a tool for correcting noisy data through resampling. Semi-variogram analysis is a robust and flexible tool for investigating spatial variability in data sets, and for assessing noise/uncertainty. It has been used by the PI in published investigations into seabed variability (Goff et al., 2002; 2004). We will apply this tool to the SEABED data bases where sufficient data density exist, investigating the variability structure (primarily rms variability, characteristic horizontal scale, and fractal dimension) as a function of environmental settings. Preliminary results indicate that the word-based estimates of mean grain size are noisier than analytic estimates, but otherwise produce accurate estimates of seabed variability (Figure 1). Through semi-variogram comparisons of word- and analytic-based mean grain size measurements, we can readily estimate and make corrections for the differences between the two measurement types.
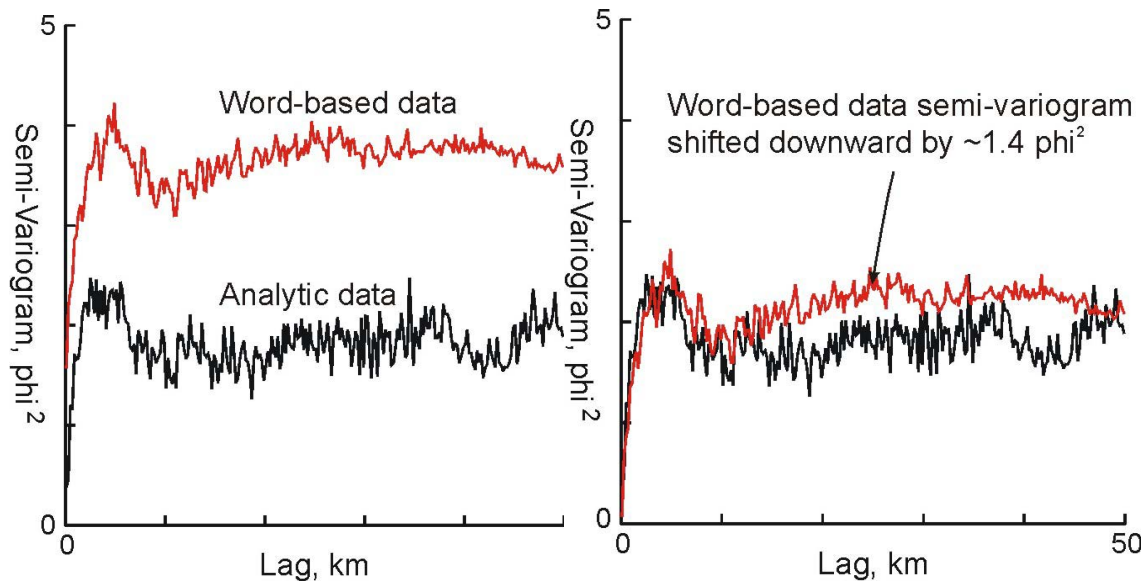

Figure 1. Semi-variograms of mean grain size measurements from the usSEABED data base for the US east coast margin, from Cape Hatteras to Long Island (left panel). Values based on analytic measurements (e.g., sieving, settling tubes; ~3000 points) were analyzed separately from those values that were derived by conversion of word-based descriptions (~10000 points). A simple static shift (right panel) brings the two curves into very good alignment, indicating that the noise (or filter) difference between them is white

Correlating seabed variability to environmental parameters will constitute one of our most significant challenges, and it is in this arena that collaboration with the USGS promises to be of critical importance. So little is presently known in this area of investigation that much of our proposed work will be exploratory. We will, in particular, investigate the predictability of variability structure. In other words: what easily measured environmental parameters (e.g., bottom wave-climate, water depth,

siliciclastic vs. carbonate) or geologic conditions (e.g., passive margin vs active margin, high sediment input vs. low, estuarine vs. open marine, etc.) can be used to constrain variability structure where samples are few or none? The null hypothesis is that there is no predictability; i.e., that every area we examine has unique variability structure uncorrelated to any environmental or geologic parameter. We doubt strongly that this is the case, however. For example, a preliminary investigation of variability on the US northeast margin (Figure 2) suggests that water depth is a strong factor in controlling variability, with greater grain size variability and shorter decorrelation scales in shallow water where oceanographic processes are more intense. Our investigations will naturally focus on those areas where sampling is dense which, at present, primarily includes the margins of the United States, Australia and Europe. Work continues to extend the resolution and coverage of the databases but already, the available data covers a very broad range of environments that can be used to test predictability hypotheses.
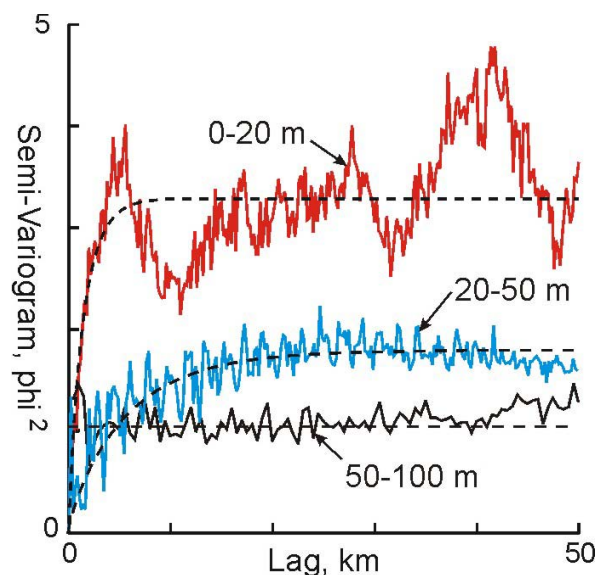

Figure 2. Semi-variograms, computed for different water depth ranges, derived from mean grain size measurements in the usSEABED data base for the US east coast margin from Cape Hatteras to Long Island.

In any geologic application, and particularly word-based mean grain size values, noisy data are sources of consternation for researchers, inhibiting interpretability and marring images with unsightly and unrealistic artifacts. Filtering is the typical solution to dealing with noisy data. However, filtering commonly suffers from ad hoc (i.e., uncalibrated, ungoverned) application, which runs the risk of erasing high variability components of the field in addition to the noise components. For this project we will establish an alternative to filtering: a methodology for correcting noise in data by finding the "best" value given the data value, its uncertainty, and the data values and uncertainties at proximal locations. The motivating rationale is that data points that are close to each other in space cannot differ by "too much", where how much is "too much" is governed by the field correlation properties. Data with large uncertainties will frequently violate this condition, and in such cases need to be corrected, or "resampled." The best solution for resampling is determined by the maximum of the likelihood

function defined by the intersection of two probability density functions (pdfs): (1) the sample pdf, with mean and variance determined by the data value and square uncertainty, respectively, and (2) the conditional pdf, whose mean and variance are determined by the kriging algorithm applied to proximal data values. A Monte Carlo sampling of the data probability space eliminates non-uniqueness, and weights the solution toward data values with lower uncertainties.

WORK COMPLETED

The primary accomplishment of the PI thus far has been to complete the development of the maximum likelihood resampling algorithm described above. This is the topic of a paper in preparation. Tests with synthetic sampling of a known field demonstrate quantitatively and qualitatively the improvement provided by this algorithm. Comparison with filtered fields demonstrates that maximum likelihood resampling does a better job at preserving the spatial statistical character of the field. Here we present two data applications of resampling:

(1) three generations of bathymetric data on the New Jersey shelf with disparate data uncertainties; and

(2) mean grain size data from the Adriatic Sea, which is combination of both analytic (low uncertainty) and word-based (higher uncertainty) sources.

The region of the US Atlantic margin chosen for analysis contains data from three different sources [*Calder*, in press]: lead-line data collected in the 1930's, echo-sounding values collected in the 1970's (both contained in the National Geophysical Data Center archives), and multibeam data collected in 1996 [*Goff et al.*, 1999]. Regions not constrained by multibeam data are marred by numerous "dimple" artifacts in the bathymetric interpolation (Figure 3). *Calder* [in press] conducted an error analysis of all three types of data and found that the lead line data were substantially biased toward shallower values; the dimples in Figure 1 are, primarily, caused by these positive errors. While *Calder* [in press] in his rendering of the bathymetry in this region chose simply to remove the lead line data in order to improve the image, here we retain them in the data set to demonstrate the utility of the maximum likelihood resampling methodology in mitigating such problems without *a priori* knowledge of their existence. Analysis of the spatial statistics of the bathymetry in this region was conducted by *Goff et al.* [1999] based on the multibeam bathymetry. The post-resampled image (Figure 4) successfully removes the dimple artifacts while leaving the multibeam data largely unmodified.

Mean grain sizes in the Adriatic sea (Figure 5) were presented earlier by *Jenkins and Goff* [submitted] in a study of optimal interpolation techniques. These data are contained in the goSEABED data base [*Jenkins* reference], and are derived from two primary sources: (1) analytic measurements of the grain size histogram, through settling tube, sedigraph and/or dry sieve techniques, and (2) conversion of word-based descriptions of bottom samples (g.e., gravel, sand, mud, silt, clay, muddy sand, silty clay, etc.) into quantitative estimates of mean grain size by applying fuzzy logic techniques. The word-based data contain, understandably, significant uncertainties compared with the analytic data [*Jenkins and Goff*, submitted]. In the interpolated data set (Figure 5), both positive and negative dimples, where data values are incompatible with nearby data points, are common. Nevertheless, the word-based mean grain size values constitute the vast majority of data values in the SEABED data bases; in the Adriatic in particular, there are less than 200 analytically derived mean grain size values versus more than 2000 word-based values. The word-based values cannot, therefore, simply be excluded without severely compromising coverage. Estimates of both the uncertainty in mean grain size data values and the semi-variogram structure of the Adriatic data set are presented in *Jenkins and Goff* [submitted].

The post-resampled image (Figure 6) shows these artifacts to be mostly removed. The post-resampled image provides a more realistic and satisfactory presentation of the data.
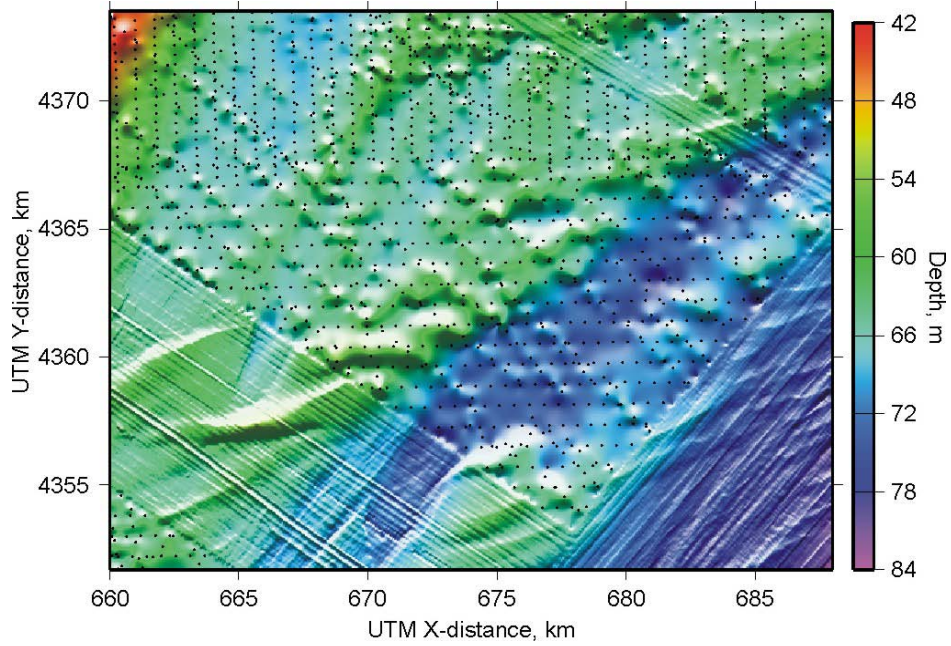

Figure 3. Region of New Jersey shelf bathymetric data chosen for application of the maximum likelihood resampling algorithm, color contoured and artificially illuminated from the north. Striated regions are areas of multibeam bathymetry data [Goff et al., 1999]. Dots indicate locations of archival lead-line and sounding data points from the NGDC outside of multibeam coverage, interpolated with a spline-in-tension algorithm [Smith and Wessel, 1990].
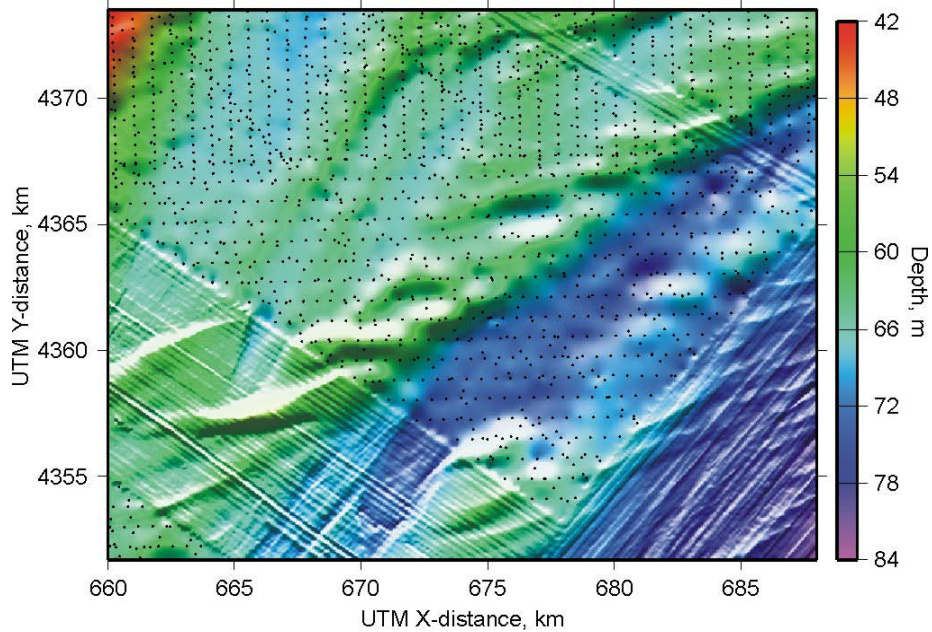

Figure 4. New Jesersey bathymetric data from Figure 3 after application of maximum likelihood resampling algorithm.
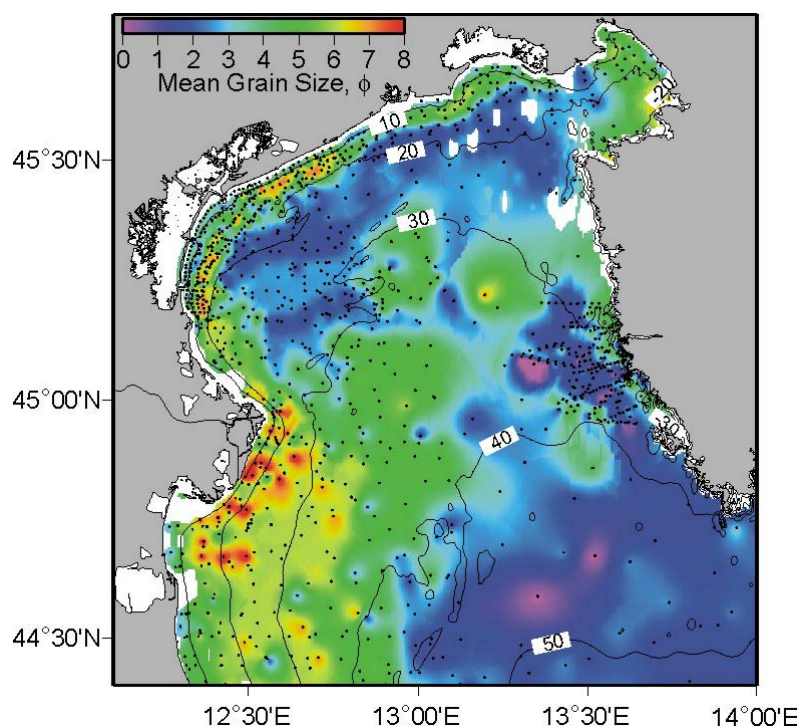
Figure 5. Mean grain sizes (in ϕ values, where grain size in mm = $2^{-\phi}$) in the northern Adriatic Sea [Jenkins and Goff, submitted]. Bathymetric contours in meters are also shown. Interpolation is accomplished through a modified version of the kriging algorithm.
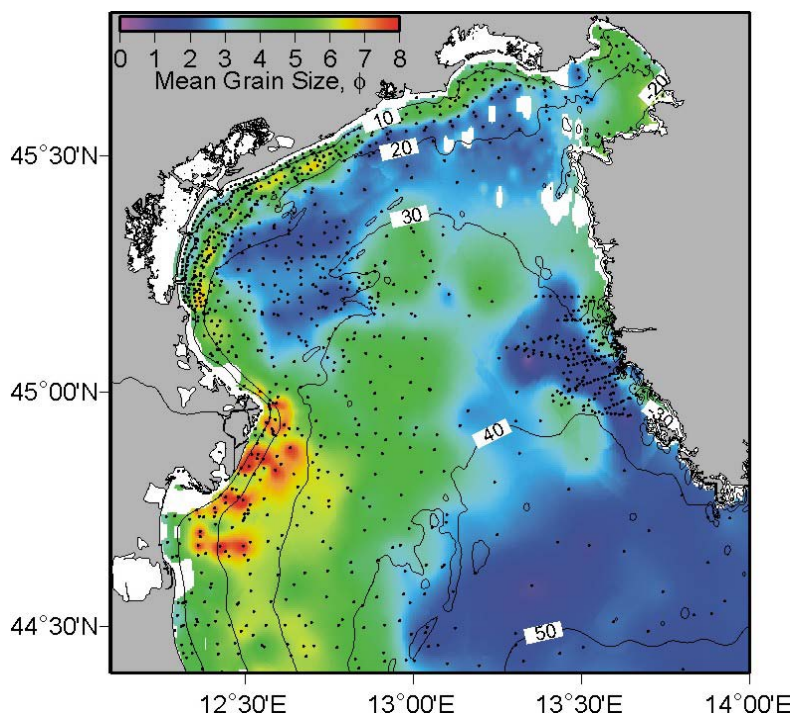

Figure 6. Northern Adriatic Sea mean grain size data from Figure 5 after application of maximum likelihood resampling algorithm.

RESULTS

The maximum likelihood resampling algorithm has proven, in both synthetic tests and disparate data applications, to be a viable method for correcting noisy data that are spatially correlated. The essential requirements for applying this method are a quantitative estimate of the uncertainty of the data and a characterization of the spatial covariance function for the sampled field. Potential applications are numerous. Maximum likelihood resampling is an important alternative to filtering. Primary advantages include: (1) an objective and optimal method for reducing noise, and (2) better preservation of the statistical properties of the sampled field. The primary disadvantage is that maximum likelihood resampling is a computationally expensive procedure. Application to large data sets will require cost/benefit considerations.

IMPACT/APPLICATIONS

This project could provide a major advance in marine science, a set of reliable methods which transform point-site seabed data into griddings that will be useful across oceanographic disciplines, sediment transport, acoustics, habitat, wave-energy generation. Our work will result in a set of software tools that will be open source, and available for inclusion any existing software packages. These tools could be of importance to the Navy, particularly in dealing with areas with sparse data, such as "denied" areas. In particular, an understanding of the relationship between environmental parameters, geologic setting and spatial variability could provide an ability to predict the amount and spatial scales of seabed variability using a parameterized semi-variogram model. This functionality provides a basis upon which to predict seabed parameters at unsampled locations, and to assess the uncertainty in that prediction. Such an understanding will have important implications for assessing acoustic prediction uncertainty. Furthermore, the semi-variogram model can be used to investigate optimal survey design, should it be possible to conduct limited sampling in denied areas via covert means (e.g., AUV's).

RELATED PROJECTS

This work is not presently linked to any other programs, but could prove useful to ONR programs such as the Ripples DRI and the Shallow Water Acoustics '06 experiment, which will make use of interpolated point data related to seabed properties.

REFERENCES

Calder, B. (in press), On the uncertainty of archive hydrographic data sets, *IEEE J. Ocean Eng*.

Goff, J. A., D. J. P. Swift, C. S. Duncan, L. A. Mayer, and J. Hughes-Clarke (1999), High resolution swath sonar investigation of sand ridge, dune and ribbon morphology in the offshore environment of the New Jersey Margin, *Mar. Geol*., *161*, 309-339.Goff, J. A.,Wheatcroft, R. A., Lee, H., Drake, D. E., Swift, D. J. P., and Fan, S. (2002), Spatial Variability of Shelf Sediments in the STRATAFORM Natural Laboratory, Northern California, *Cont. Shelf Res*., 22, 1199-1223.

Goff, J. A., B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord (2004), Seabed characterization on the New Jersey middle and outer shelf: Correlability and spatial variability of seafloor sediment properties, *Mar. Geol*., *209*, 147-172.

Smith, W. H. F., and P. Wessel, P. (1990), Gridding with continuous curvature splines in tension: Geophysics, v. 55, no. 3, p. 293-305.

PUBLICATIONS

Jenkins, C. J., and J. A. Goff (submitted), Competent interpolation for seabed substrates, with uncertainty calculations, *IEEE J. Ocean Eng.*